35.G2761

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	:	Examiner: UNASSIGNED
YASUO OKUTANI, ET AL.	)	
	:	Group Art Unit: 2641
Application No.: 09/818,607	)	
	:	
Filed: March 28, 2001	)	
	:	
For: SPEECH SIGNAL PROCESSING APPARATUS AND METHOD, AND STORAGE MEDIUM	)	
	:	July 19, 2001

#3
7/27/01
MB

Commissioner for Patents
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

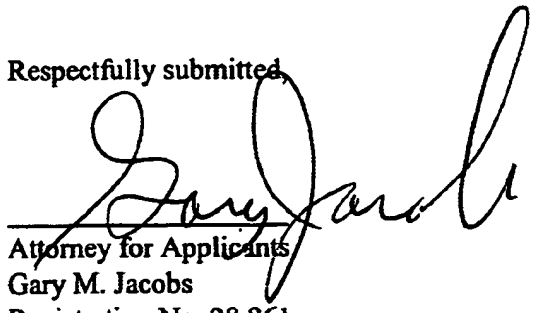
Applicants hereby claim priority under the International Convention and all rights to which they are entitled under 35 U.S.C. § 119 based upon the following Japanese Priority Application No.:

2000-099533 filed March 31, 2000.

A certified copy of the priority document is enclosed.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our below-listed address.

Respectfully submitted,



Attorney for Applicants
Gary M. Jacobs
Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile No.: (212) 218-2200

DC_MAIN 59894 v 1

DC_MAIN 59894 v 1

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２０００年 ３月３１日

出願番号
Application Number: 特願２０００－０９９５３３

出願人
Applicant(s): キヤノン株式会社

Inventor: Yasuo Okutani, et al.
Appln. No: 09/818,607
Filed: 3/31/01

２００１年 ４月２０日

特許庁長官
Commissioner,
Japan Patent Office 及川耕造

出証番号 出証特２００１－３０３３１６０

【書類名】　特許願
【整理番号】　4172019
【提出日】　平成12年 3月31日
【あて先】　特許庁長官殿
【国際特許分類】　G01L　5/04
【発明の名称】　音声情報処理装置及びその方法と記憶媒体
【請求項の数】　19
【発明者】
　【住所又は居所】　東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内
　【氏名】　奥谷　泰夫
【発明者】
　【住所又は居所】　東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内
　【氏名】　小森　康弘
【特許出願人】
　【識別番号】　000001007
　【氏名又は名称】　キヤノン株式会社
【代理人】
　【識別番号】　100076428
　【弁理士】
　【氏名又は名称】　大塚　康徳
　【電話番号】　03-5276-3241
【選任した代理人】
　【識別番号】　100101306
　【弁理士】
　【氏名又は名称】　丸山　幸雄
　【電話番号】　03-5276-3241

【選任した代理人】

【識別番号】　100115071

【弁理士】

【氏名又は名称】　大塚　康弘

【電話番号】　03-5276-3241

【手数料の表示】

【予納台帳番号】　003458

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　0001010

【プルーフの要否】　要

【書類名】　　明細書

【発明の名称】　　音声情報処理装置及びその方法と記憶媒体

【特許請求の範囲】

【請求項１】　音声素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

前記歪出力手段から出力された歪に基づいて、音声合成に使用する音声素片を選択する選択手段と、

前記選択手段により選択された音声素片に基づいて、音声合成を行う音声合成手段と、

を有することを特徴とする音声情報処理装置。

【請求項２】　前記選択手段は、歪が最小となるように音声合成に使用する音声素片を選択することを特徴とする請求項１に記載の音声情報処理装置。

【請求項３】　前記歪出力手段は、前記音素素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項１又は２に記載の音声情報処理装置。

【請求項４】　前記歪出力手段は、前記接続歪と、前記変形歪との重み付き和として前記歪を算出することを特徴とする請求項３に記載の音声情報処理装置。

【請求項５】　前記歪出力手段は、ケプストラム距離を用いて前記接続歪を算出することを特徴とする請求項３又は４に記載の音声情報処理装置。

【請求項６】　前記歪出力手段は、ケプストラム距離を用いて前記変形歪を算出することを特徴とする請求項３又は４に記載の音声情報処理装置。

【請求項７】　前記歪出力手段は、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項３又は４に記載の音声情報処理装置。

【請求項８】　前記歪出力手段は、前記接続歪を記憶したテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項３又は４に記載の音声情報処理装置。

【請求項９】　テキストデータを入力する入力手段と、

前記テキストデータを言語解析する言語解析手段と、

前記言語解析手段により解析された結果に基づいて前記所定の韻律情報を生成する韻律情報生成手段と、

を更に備えることを特徴とする請求項１乃至８のいずれか１項に記載の音声情報処理装置。

【請求項１０】　音声素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

前記歪出力工程で出力された歪に基づいて、音声合成に使用する音声素片を選択する選択工程と、

前記選択工程で選択された音声素片に基づいて、音声合成を行う音声合成工程と、

を有することを特徴とする音声情報処理方法。

【請求項１１】　前記選択工程では、前記歪が最小となるように音声合成に使用する音声素片を選択することを特徴とする請求項１０に記載の音声情報処理方法。

【請求項１２】　前記歪出力工程では、前記音素素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項１０又は１１に記載の音声情報処理方法。

【請求項１３】　前記歪出力工程では、前記接続歪と、前記変形歪との重み付き和として前記歪を算出することを特徴とする請求項１２に記載の音声情報処理方法。

【請求項１４】　前記歪出力工程では、ケプストラム距離を用いて前記接続歪を算出することを特徴とする請求項１２又は１３に記載の音声情報処理方法。

【請求項１５】　前記歪出力工程では、ケプストラム距離を用いて前記変形歪を算出することを特徴とする請求項１２又は１３に記載の音声情報処理方法。

【請求項１６】　前記歪出力工程では、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項１２

又は１３に記載の音声情報処理方法。

【請求項１７】　前記歪算出工程では、前記接続歪を記憶したテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項１２又は１３に記載の音声情報処理方法。

【請求項１８】　テキストデータを入力する入力工程と、

前記テキストデータを言語解析する言語解析工程と、

前記言語解析工程で解析された結果に基づいて前記所定の韻律情報を生成する韻律情報生成工程と、

を更に備えることを特徴とする請求項１０乃至１７のいずれか１項に記載の音声情報処理方法。

【請求項１９】　請求項１０乃至１８のいずれか１項に記載の方法を実行するプログラムを記憶したことを特徴とする、コンピュータにより読取り可能な記憶媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、音声素片を編集、接続して音声合成を行なう音声情報処理装置及びその方法と、その方法を実現するプログラムを記憶した記憶媒体に関するものである。

【０００２】

【従来の技術】

近年、テキストデータを入力し、そのテキストデータを言語解析してポーズ部分、無音時間の長さ、アクセントの生成などを行なって韻律情報を生成し、更に、その韻律情報に従って音声素片を記憶している素片辞書を検索し、対応する音声素片を読み出して音声合成する音声合成装置が知られている。

【０００３】

このような音声合成装置では、その読み出した音声素片を１ピッチ波形単位で複製、削除しながら、所望のピッチ間隔で貼り合わせて編集し（PSOLA：ピッチ同期波形重畳法）、それらの音声素片を接続する音声合成方式が主流となってい

る。

【０００４】

【発明が解決しようとする課題】

このような技術を利用して合成された音声には、音声素片を編集（変形）することによる歪（以下、変形歪）と、音声素片同士を接続することによって生じる歪（以下、接続歪）とが含まれ、これら２つの歪が合成音声の品質劣化を引き起こす大きな要因となっている。

【０００５】

本発明は上記従来例に鑑みてなされたもので、接続や変形に基づく歪の影響を小さくする音声情報処理装置及びその方法と、その方法を実現するプログラムを記憶した記憶媒体を提供することを目的とする。

【０００６】

【課題を解決するための手段】

上記目的を達成するために本発明の音声情報処理装置は以下のような構成を備える。即ち、

音声素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

前記歪出力手段から出力された歪に基づいて、音声合成に使用する音声素片を選択する選択手段と、

前記選択手段により選択された音声素片に基づいて、音声合成を行う音声合成手段と、

を有することを特徴とする。

【０００７】

上記目的を達成するために本発明の音声情報処理方法は以下のような工程を備える。即ち、

音声素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

前記歪出力工程で出力された歪に基づいて、音声合成に使用する音声素片を選択する選択工程と、

前記選択工程で選択された音声素片に基づいて、音声合成を行う音声合成工程と、

を有することを特徴とする。

【０００８】

【発明の実施の形態】

以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。

【０００９】

［実施の形態１］

図１は、本発明の実施の形態に係る音声合成装置のハードウェア構成を示すブロック図である。尚、本実施の形態では、一般的なパーソナルコンピュータを音声合成装置として用いる場合について説明するが、本発明は専用の音声合成装置であっても、また他の形態の装置であっても良い。

【００１０】

図１において、１０１は制御メモリ（ＲＯＭ）で、中央処理装置（ＣＰＵ）１０２で使用される各種制御データを記憶している。ＣＰＵ１０２は、ＲＡＭ１０３に記憶された制御プログラムを実行して、この装置全体の動作を制御している。１０３はメモリ（ＲＡＭ）で、ＣＰＵ１０２による各種制御処理の実行時、ワークエリアとして使用されて各種データを一時的に保存するとともに、ＣＰＵ１０２による各種処理の実行時、外部記憶装置１０４から制御プログラムをロードして記憶している。この外部記憶装置は、例えばハードディスク、ＣＤ－ＲＯＭ等を含んでいる。１０５はＤ／Ａ変換器で、音声信号を示すデジタルデータが入力されると、これをアナログ信号に変換してスピーカ１０９に出力して音声を再生する。１０６は入力部で、オペレータにより操作される、例えばキーボードや、マウス等のポインティングデバイスを備えている。１０７は表示部で、例えばＣＲＴや液晶等の表示器を有している。１０８はバスで、これら各部を接続している。１１０は音声合成ユニットである。

【００１１】

以上の構成において、本実施の形態の音声合成ユニット１１０を制御するための制御プログラムは外部記憶装置１０４からロードされてＲＡＭ１０３に記憶さ

れ、その制御プログラムで用いる各種データは制御メモリ１０１に記憶されている。これらのデータは、中央処理装置１０２の制御の下にバス１０８を通じて適宜メモリ１０３に取り込まれ、中央処理装置１０２による制御処理で使用される。Ｄ／Ａ変換器１０５は、制御プログラムを実行することによって作成される音声波形データ（ディジタル信号）をアナログ信号に変換してスピーカ１０９に出力する。

【００１２】

図２は、本実施の形態に係る音声合成ユニット１１０の構成を示すブロック図である。

【００１３】

図２において、２０１は入力部１０６や外部記憶装置１０４から任意のテキストデータを入力するテキスト入力部、２０２は解析辞書、２０３は言語解析部、２０４は韻律生成規則保持部、２０５は韻律生成部、２０６は素片辞書である音声素片保持部、２０７は音声素片選択部、２０８は音声素片編集・接続部、２０９は音声波形出力部である。

【００１４】

以上の構成において、言語解析部２０３が、解析辞書２０２を参照して、テキスト入力部２０１から入力されるテキストの言語解析を行なう。こうして解析された結果が韻律生成部２０５に入力される。韻律生成部２０５は、言語解析部２０３における解析結果と、韻律生成規則保持部２０４に保持されている韻律生成規則に関する情報とを基に音韻系列と韻律情報を生成して音声素片選択部２０７及び音声素片編集・接続部２０８に出力する。続いて、音声素片選択部２０７は、韻律生成部２０５から入力される韻律生成結果を用いて、音声素片保持部２０６に保持されている音声素片から対応する音声素片を選択する。音声素片編集・接続部２０８は、韻律生成部２０５から入力される韻律生成結果に従って、音声素片選択部２０７から出力される音声素片を編集及び接続して音声波形を生成する。こうして生成された音声波形は、音声波形出力部２０９で出力される。

【００１５】

次に、上記構成を備えた本実施の形態の音声合成処理について説明する。

【００１６】

図３は、本実施の形態に係る音声合成ユニット１１０における音声合成処理の流れを示すフローチャートである。

【００１７】

まずステップＳ３０１で、テキスト入力部２０１は、文、文節、単語等の単位毎に、テキストデータを入力してステップＳ３０２に移る。ステップＳ３０２では、言語解析部２０３により当該テキストデータの言語解析を行う。次にステップＳ３０３に進み、音韻生成部２０５は、ステップＳ３０２で解析された結果と所定の韻律規則とに基づいて、音韻系列と韻律情報を生成する。次にステップＳ３０４に進み、各音韻毎に、ステップＳ３０３で得られた韻律情報と所定の音韻環境とに基づいて、音声素片選択部２０７が音声素片保持部２０６に登録されている音声素片を選択する。次にステップＳ３０５に進み、その選択された音声素片及びステップＳ３０３で生成された韻律情報とに基づいて、音声素片編集・接続部２０８により音声素片の編集および接続を行なってステップＳ３０６に進む。ステップＳ３０６では、音声素片編集・接続部２０８によって生成された音声波形を、音声波形出力部２０９が音声信号として出力する。このようにして、入力されたテキストに対応する音声が出力されることになる。

【００１８】

図４は、図３のステップＳ３０４（音声素片選択）の処理の詳細を示すフローチャートである。

【００１９】

このステップＳ３０４では、音声素片同士の接続歪（後述する）と、音声素片の変形歪（後述する）とに基づいて決定される歪値（後述）に従って、動的計画法により、入力テキストデータの全体に亙って歪値が最小となる音声素片系列を決定する。つまり、韻律生成部２０５が生成する音韻系列Ｐｎ（０≦ｎ＜Ｎ）の先頭（ｎ＝０）から順に処理することになる。まず最初のｎ＝０にセットし、ステップＳ４０１で、音韻系列の終端まで処理が終了していない場合、つまりｎ＜Ｎの場合はステップＳ４０２に進み、ｎ番目の音韻における音声素片の候補を音声素片保持部２０６から取り出し、それら音声素片の候補の個数をMnとしてス

テップＳ４０３に進む。ステップＳ４０３では、まず最初にｍ＝０にセットし、上記ｎ，ｍで特定される音声素片候補Ｐｎ,ｍ（０≦ｍ＜Ｍｎ）に着目し、ｎ番目の音韻における音声素片の候補の先頭（ｍ＝０）から順に処理すして、候補の最後まで処理が終了していない場合、つまりｍ＜Ｍｎの場合はステップＳ４０４に進むが、最後まで処理が終了した場合は、次の音韻の処理に移行するためｎ＝ｎ＋１としてステップＳ４０１に戻る。ステップＳ４０４では、一つ前の音韻Ｐｎ-1の各音声素片候補Ｐｎ-1,ｋ（０≦ｋ＜Ｍｎ-1：Ｍｎ-1は一つ前の音韻Ｐｎ-1の音声素片候補の数）と、候補Ｐｎ,ｍとの間の歪値Ｄｋ,ｍをそれぞれ計算してステップＳ４０５に進む。ステップＳ４０５では、候補Ｐｎ,ｍに至るまでの歪値の総和の最小値である総和Ｓｎ,ｍを求める。この総和Ｓｎ,ｍは、次式で表現される。

【００２０】

$$Sn,m = \min(Sn-1,k + Dk,m) \quad : 0 \leqq k < Mn-1$$

この式において、ｍｉｎ（）は、ｋを“０”から“Ｍｎ-1”まで変化させた場合の最小値を意味する。また、そのときのｋの値をＰＲＥｎ,ｍとして保持しておく。このＰＲＥｎ,ｍは、候補Ｐｎ,ｍに至るまでの歪値の総和が最小となる経路を示し、ステップＳ４０６において最小歪経路を特定するために利用される。この候補Ｐｎ,ｍの総和Ｓｎ,ｍとＰＲＥｎ,ｍが決定したら、次の音声素片候補に対する処理を行なうためにｍ＝ｍ＋１としてステップＳ４０３に戻る。

【００２１】

こうしてステップＳ４０１で、最終であるＮ番目の音韻系列までの処理が終了するとステップＳ４０６に進み、総和ＳＮ-1,ｍ（０≦ｍ＜Ｍｎ）が最小となる候補ＰＮ-1,ｍを特定し、そこから順次ＰＲＥｎ,ｍを辿ることによって最小歪経路となる音声素片系列を特定する。こうして音声素片系列が特定されたら図３のステップＳ３０５に進んで、これら特定された音声素片の編集・接続を実行する。

【００２２】

図５は、ｎ番目の音韻（現在注目している音韻）の音声素片候補Ｐｎ,1における総和Ｓｎ,1の算出を模式的に示した図である。本実施の形態では、音韻の単位としてdiphoneを採用した場合について記述する。

【００２３】

図中、一つの円が音声素片の１候補Ｐn,mを示し、円内の数字が歪値の総和の最小値である総和Ｓn,mを示している。また矢印は、上述のＰＲＥn,mを指す。また、四角で囲まれた数字は、音声素片候補Ｐn,mの歪値Ｄk,mを表わしている。

【００２４】

次に、本実施の形態における歪値について説明する。

【００２５】

ここでは、歪値Ｄk,mを接続歪Ｄcと変形歪Ｄtの重み付き和として定義する。即ち、

Ｄ ＝ ｗ×Ｄc ＋ （１－ｗ）×Ｄt ：（０≦ｗ≦１）

ここで、重み係数ｗは、予備実験など経験的に求める係数で、ｗ＝０の場合は、歪値が変形歪Ｄtのみで説明され、ｗ＝１の場合は歪値が接続歪Ｄcのみに依存することになる。

【００２６】

図５では、音声素片候補Ｐn,1の一つ前の音韻の音声素片候補Ｐn-1,２５０との間の歪値Ｄ2,1が“３”であり、音声素片候補Ｐn-1,２５０に至るまでの歪値の総和Ｓn-1,2が“８”であるため、経路５１がＰＲＥn,1として決定される。

【００２７】

図６は、本実施の形態における接続歪Ｄcの求め方を説明する図である。

【００２８】

接続歪Ｄcは、一つ前の音声素片と現在の音声素片との接続箇所において生じる歪で、本実施の形態では、ケプストラム距離を用いて表す。ここでは、音声素片境界が存在するフレーム６０，６１（フレーム長５ミリ秒、分析窓幅２５．６ミリ秒）と、それを挟む前後それぞれの２フレームからなる計５フレームを接続歪の算出対象とする。ケプストラムは、０次（パワー）～１６次（パワー）までの計１７次元とする。このケプストラムベクトルの各要素の差の絶対値の和を、現在注目している音声素片における接続歪とする。一つ前の音声素片における終端部のケプストラムベクトルの各要素をＣp i,j（ｉはフレーム番号で、ｉ＝０が音声素片境界があるフレームである。ｊはベクトルの要素番号を示す）、当該音声素片における始端部のケプストラムベクトルの各要素をＣc i,jとすると、

現在注目している音声素片の接続歪Ｄｃは、

$$Dc=\Sigma\Sigma\,|\,Cp_{i,j}-Cc_{i,j}\,|$$

となる。ここで最初のΣはｉ＝－２～２の総和を示し、次のΣはｊ＝０～１６の総和を示している。

【００２９】

図７は、本実施の形態に係る変形歪Ｄｃの求め方を説明する図である。

【００３０】

ここでは、PSOLA法によりピッチ間隔を広げる場合について図示している。矢印はピッチマーク、点線は変形前と変形後のピッチ素片の対応関係を表わしている。本実施の形態では、各ピッチ素片（微細素片ともいう）の変形前後のケプストラム距離に基づいて変形歪を表す。具体的には、まず、変形後のあるピッチ素片（例えば７０で示す）のピッチマーク７１を中心にハニング窓７２（窓長２５．６ミリ秒）をかけ、そのピッチ素片７０を周辺のピッチ素片を含めて切り出す。こうして切り出したピッチ素片７０をケプストラム分析する。次に、ピッチマーク７１に対応する変形前のピッチ素片７３のピッチマーク７４を中心にして同じ窓長のハニング窓７５でピッチ素片を切り出し、変形後の場合と同様にしてケプストラムを求める。このようにして求めたケプストラム同士の距離を、着目しているピッチ素片７０の変形歪として、変形後のピッチ素片とそれに対応する変形前のピッチ素片間の変形歪の総和をPSOLAで採用されるピッチ素片数Ｎｐで割った値を、その音声素片の変形歪とする。こうして求められる変形歪を式で記述すると以下のようになる。

【００３１】

$$Dt=\Sigma\Sigma\,|\,Corg_{i,j}-Ctar_{i,j}\,|\,/\,Np$$

ここで最初のΣは、ｉ＝１からＮまでの総和を示し、次のΣはｊ＝０～１６までの総和を示している。またＣtar ｉ,ｊは、変形後のｉ番目のピッチ素片のケプストラムのｊ次元目の要素を表わし、同様に、Ｃorg ｉ,ｊは、変形後に対応する変形前のピッチ素片のケプストラムのｊ次元目の要素を表わしている。

【００３２】

このように本実施の形態１によれば、各音声素片における接続歪及び変形歪を

求め、これら歪を基に重み付け計算を行なって各音声素片における歪値を求め、この歪値の総和が最小となる音声素片系列を特定して音声合成することにより、良好な音声合成結果を得ることができるという効果がある。

【００３３】

［実施の形態２］

前述の実施の形態１では、音韻の単位としてdiphoneを用いる場合について記述したが、本発明はこれに限定されるものではなく、音素や半diphoneなどを単位としてもよい。半diphoneとは、diphoneを音素境界で２つに分割したもののことである。

【００３４】

図８は、半diphoneを単位とした場合の概念図である。この半diphoneを単位とした場合のメリットについて簡単に説明する。任意のテキストを合成する場合、素片辞書は、全種類のdiphoneを用意しておく必要がある。これに対して、半diphoneを単位とした場合は、足りない半diphoneを別の半diphoneで代替できる。例えば、半diphoneの「/a.b.0/(diphone　a.bの左側）」の代わりに「/a.n.0/」を利用しても、音質の劣化を少なくして良好に音声を再生できる。これにより、素片辞書のサイズをより小さくできる。

【００３５】

［実施の形態３］

前述の実施の形態１及び２では、音韻の単位としてdiphoneや音素や半diphoneを用いる場合について説明したが、本発明はこれに限定されるものではなく、これらを混合して用いてもよい。例えば、利用頻度が高い音韻については、diphoneを単位として、利用頻度が低い音韻については、２つの半diphoneを用いて表現するようにしても良い。

【００３６】

図９は、音声素片単位を混合した場合の一例を示した図で、ここでは音韻「o.w」がdiphoneで表され、その前後の音韻は半diphoneで表されている。

【００３７】

［実施の形態４］

実施の形態３において、元のデータベース中で連続する場所から取り出されたかどうかの情報を持ち、連続していた場合は、半diphoneの組を仮想的にdiphoneとして扱うようにしてもよい。つまり、元のデータベース中で連続するということは接続歪が“０”であるため、この場合には変形歪だけを考慮すればよいことになり計算量を大幅に軽減できる。

【００３８】

図１０は、この様子を表わした概念図である。図中の線上の数字は接続歪を表している。

【００３９】

図１０において、１１００で示される半diphoneの組は、元のデータベース中で連続する場所から取り出されたものであり、その接続歪みは“０”に一義的に決定されている。また１１０１で示された半diphoneの組は、元のデータベース中で連続する場所から取り出されたものではないため、それぞれに対して接続歪みが計算される。

【００４０】

［実施の形態５］

上述の実施の形態１では、動的計画法を、１単位のテキストデータから得られた音韻系列全体に対して適用する場合について説明したが、本発明はこれに限定されるものではない。例えば、ポーズや無音部分までを一つの区間として音韻系列を分割し、各区間毎に動的計画法を実行してもよい。尚、ここで言う無音部分とは、p,t,kなどの無音部分のことである。このようなポーズや無音部分では、接続歪が“０”であると考えられるため、このような分割が有効となる。これにより、区間ごとに適当な選択結果を得ることができるだけでなく、合成音声の生成に要する時間が短縮できる。

【００４１】

［実施の形態６］

前述の実施の形態１では、接続歪の計算にケプストラムを用いる場合について説明したが、本発明はこれに限定されるものではなくい。例えば、接続点の前後に亙る波形の差分の和を用いて接続歪を求めてもよい。またスペクトル距離など

を用いて接続歪を求めてもよい。この場合、接続点はピッチマークに同期させるのが、より好ましい。

【００４２】

［実施の形態７］

前述の実施の形態１では、接続歪の計算において、窓長、シフト長、ケプストラムの次数、フレーム数などを具体的数字を使って説明したが、本発明はこれに限定されるものではない。任意の窓長、シフト長、次数、フレーム数を使って接続歪を算出してもよい。

【００４３】

［実施の形態８］

前述の実施の形態１では、接続歪の計算にケプストラムの次数ごとに差分を取ったものの総和を用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、各次数を統計的性質などを使って正規化（正規化係数ｒｊ）してもよい。この場合の接続歪Ｄｃは、

$$Dc=\Sigma\Sigma\ (r_j\times|Cpre_{i,j}-Ccur_{i,j}|)$$

となる。ここで、最初のΣはｉ＝－２～２の総和を、次のΣはｊ＝０～１６までの総和を示す。

【００４４】

［実施の形態９］

実施の形態１では、ケプストラムの次数ごとの差分の絶対値をベースに接続歪の算出を行なう場合について説明したが、本発明はこれに限定されるものではない。例えば、差分の絶対値の累乗（累数が偶数の場合は絶対値でなくてもよい）をベースに接続歪の算出を行なってもよい。ここで累数をＮとすると、接続歪Ｄｃは、

$$Dc=\Sigma\Sigma\ |Cpre_{i,j}-Ccur_{i,j}|\hat{}N$$

となる。ここで“^Ｎ”はＮ乗を示す。ここでＮの値を大きくすることは、大きな差分について敏感になることを意味しているので、その結果、接続歪が平均的に小さくなるように働くことになる。

【００４５】

［実施の形態１０］

前述の実施の形態１では、変形歪としてケプストラムを用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、変形前後の一定区間の波形の差分の和を用いて変形歪を求めてもよい。また、スペクトル距離などを用いて変形歪としてもよい。

【００４６】

［実施の形態１１］

前述の実施の形態１では、変形歪を波形から得られる情報を基に算出する場合について説明したが、本発明はこれに限定されるものではない。例えば、PSOLAによるピッチ素片の削除および複製の回数などを変形歪を算出する要素としてもよい。

【００４７】

［実施の形態１２］

前述の実施の形態１では、音声合成時に音声素片を読み出す毎に接続歪を計算する場合について説明したが、本発明はこれに限定されるものではない。例えば、接続歪を予め計算しておき、テーブルとして保持しておいてもよい。

【００４８】

図１１は、diphone「/a.r/」とdiphone「/r.i/」との間の接続歪を記憶したテーブルの一例を示す図である。ここでは縦軸に「/a.r/」の音声素片、横軸に「/r.i/」の音声素片をとっている。例えば、「/a.r/」の「id3」の音声素片と「/r.i/」の「id2」の音声素片との接続歪は“３．６”で表されている。このように接続可能なdiphone間の接続歪を全てテーブル化して用意することにより、音声素片同士の合成時の接続歪の算出がテーブルの参照だけで済むため、その計算量を大幅に軽減でき、算出時間を大幅に短縮できる。

【００４９】

［実施の形態１３］

前述の実施の形態１では、音声合成時に、音声素片編集する毎に変形歪を計算する場合について説明したが、本発明はこれに限定されるものではない。例えば、変形歪を予め計算しておき、テーブルとして保持しておいてもよい。

【００５０】

図１２は、あるdiphoneを基本周波数と音韻時間長について変化させた場合の変形歪をテーブルで表した図である。

【００５１】

図中、μは、そのdiphoneの統計的な平均値を示し、σは標準偏差である。具体的な表の作成方法としては、次のような作成方法が考えられる。まず、基本周波数と音韻時間長に関して統計的に平均値と分散を求める。次に、それらを基に（５×５＝）２５通りの基本周波数と音韻時間長をターゲットとしてPSOLA法をそれぞれ適用し、テーブルの変形歪を一つずつ求めていけばよい。合成時は、ターゲットの基本周波数と音韻時間長が決まれば、テーブルの近傍の値で内挿（もしくは外挿）することによって、変形歪を推定することが可能である。

【００５２】

図１３は、合成時に変形歪を推定するための具体例を示した図である。

【００５３】

図中、黒丸がターゲットの基本周波数と音韻時間長であり、このとき、各格子点の変形歪がテーブルからＡ，Ｂ，Ｃ，Ｄと求まっていると仮定すると、変形歪Ｄｔは、以下の式により求めることができる。

$$Dt=\{A\cdot(1-y)+C\cdot y\}\times(1-x)+\{B\cdot(1-y)+D\cdot y\}\times x$$

【００５４】

［実施の形態１４］

前述の実施の形態１３では、変形歪テーブルの格子点として、そのdiphoneの統計的な平均値と標準偏差を基に５×５のテーブルを作成したが、本発明はこれに限定されるものではなく、任意の格子点を持つテーブルとしてもよい。また、格子点を平均値などに依らず決定的に与えてもよいものとする。例えば、韻律推定で推定されうる範囲を等分割するなどもよいものとする。

【００５５】

［実施の形態１５］

前述の実施の形態１では、接続歪と変形歪の重み和で歪を定量化する場合について説明したが本発明はこれに限定されるものではなく、接続歪と変形歪それぞ

れに閾値を設定しておき、どちらか一方でもその閾値を越えた場合はその音声素片が選択されないようにして、十分大きな歪の値を与えるようにしてもよい。

【００５６】

上記実施の形態においては、各部を同一の計算機上で構成する場合について説明したが本発明はこれに限定されるものではなく、例えばネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【００５７】

上記実施の形態においては、各部を同一の計算機上で構成する場合について説明したが、これに限定されるものではなく、ネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【００５８】

上記実施の形態においては、プログラムを制御メモリ（ＲＯＭ）に保持する場合について説明したが、これに限定されるものではなく、外部記憶など任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

【００５９】

なお本発明は、複数の機器から構成されるシステムに適用しても、１つの機器からなる装置に適用してもよい。前述した実施の形態の機能を実現するソフトウエアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記録媒体に格納されたプログラムコードを読み出し実行することによっても達成される。

【００６０】

この場合、記録媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。このようなプログラムコードを供給するための記録媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、磁気テープ、不揮発性のメモリカード、ＲＯＭなどを用いることができる。

【００６１】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているＯＳなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【００６２】

更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【００６３】

以上説明したように本実施の形態によれば、音声合成に際して、音声素片を選択する際に接続歪と変形歪を基準とするようにしたため、音質の劣化を最小限に抑えた音声素片系列を求めて音声合成できる。

【００６４】

【発明の効果】

以上説明したように本発明によれば、接続や変形に基づく歪の影響を小さくした音声を合成することができる。

【図面の簡単な説明】

【図１】

本発明の実施の形態に係る音声合成装置のハードウェア構成を示すブロック図である。

【図２】

本発明の実施の形態１に係る音声合成装置の機能構成を示すブロック図である。

【図３】

本実施の形態に係る音声合成装置における処理の流れを示すフローチャートである。

【図４】

図３のステップＳ３０４の音素素片選択処理の詳細を示すフローチャートである。

【図５】

ｎ番目の音韻の音声素片候補Ｐｎ,1における最小歪の総和Ｓｎ,1の算出を模式的に示した図である。

【図６】

本発明の実施の形態に係る音声素片の接続歪を説明する図である。

【図７】

本発明の実施の形態に係る音声素片の変形を説明する図である。

【図８】

半diphoneを単位とした場合の概念図である。

【図９】

本発明の実施の形態３に係る音声素片の単位をdiphoneと半diphoneとで混合した場合を説明する図である。

【図１０】

本発明の実施の形態４に係る音声素片の単位を取り出した半diphoneによって混合した例を示した図である。

【図１１】

本発明の実施の形態１２に係るdiphoneの /a.r/ と/r.i/間の接続歪を決定するテーブル構成例を示す図である。

【図１２】

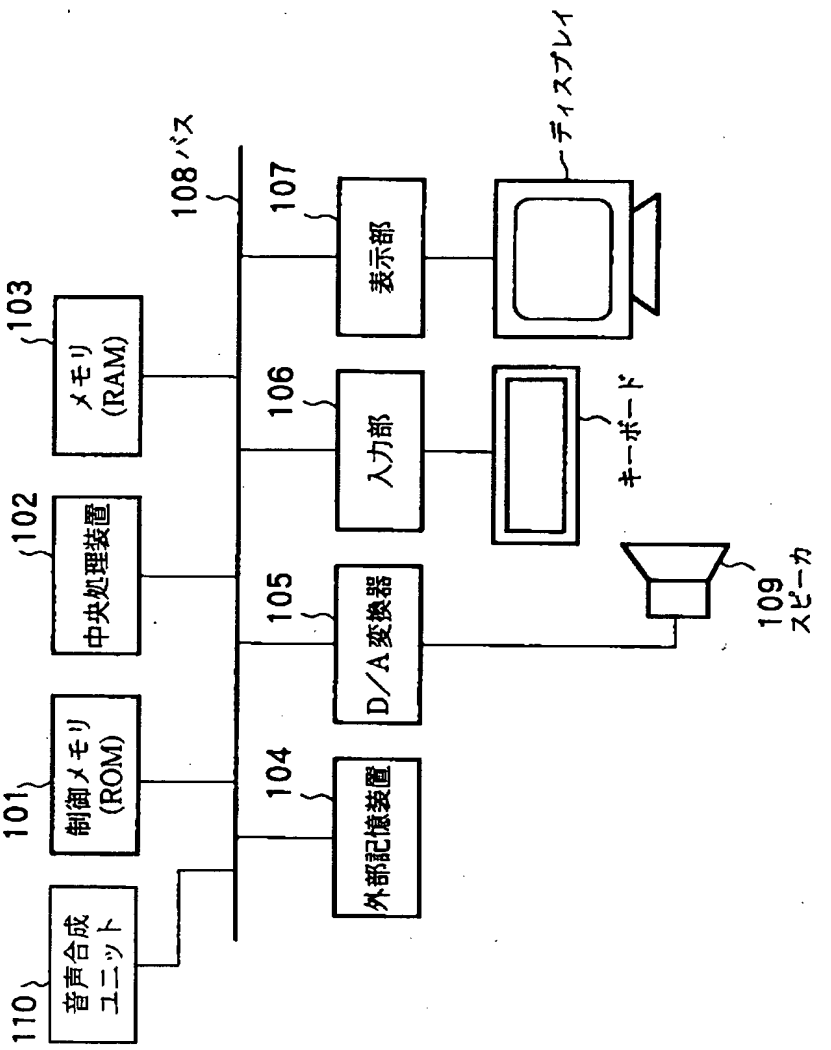
本発明の実施の形態１３に係る変形歪を表わすテーブル例を示す図である。

【図１３】

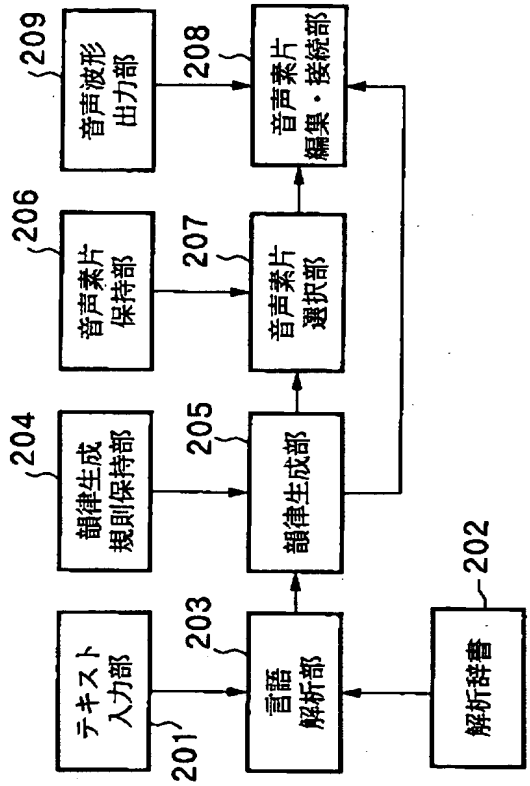
本発明の実施の形態１３に係る変形歪を推定する具体例を示した図である。

【書類名】　図面

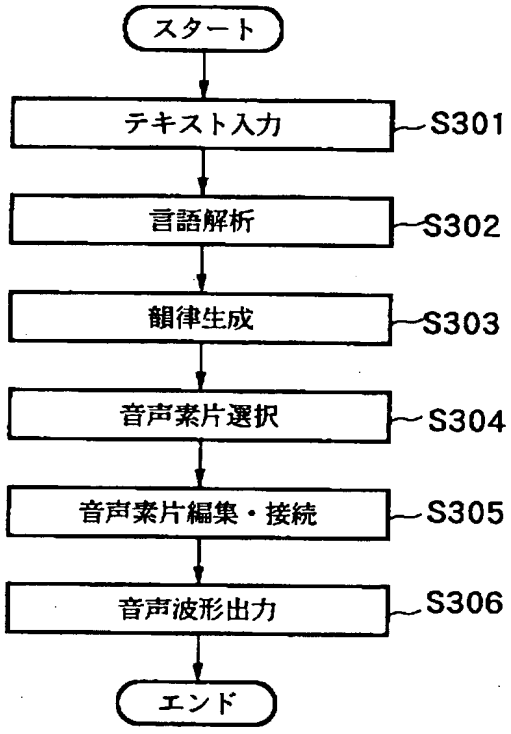
【図１】


108 バス
107
表示部
ディスプレイ
103
メモリ (RAM)
106
入力部
キーボード
102
中央処理装置
105
D／A変換器
109
スピーカ
101
制御メモリ (ROM)
104
外部記憶装置
110
音声合成ユニット

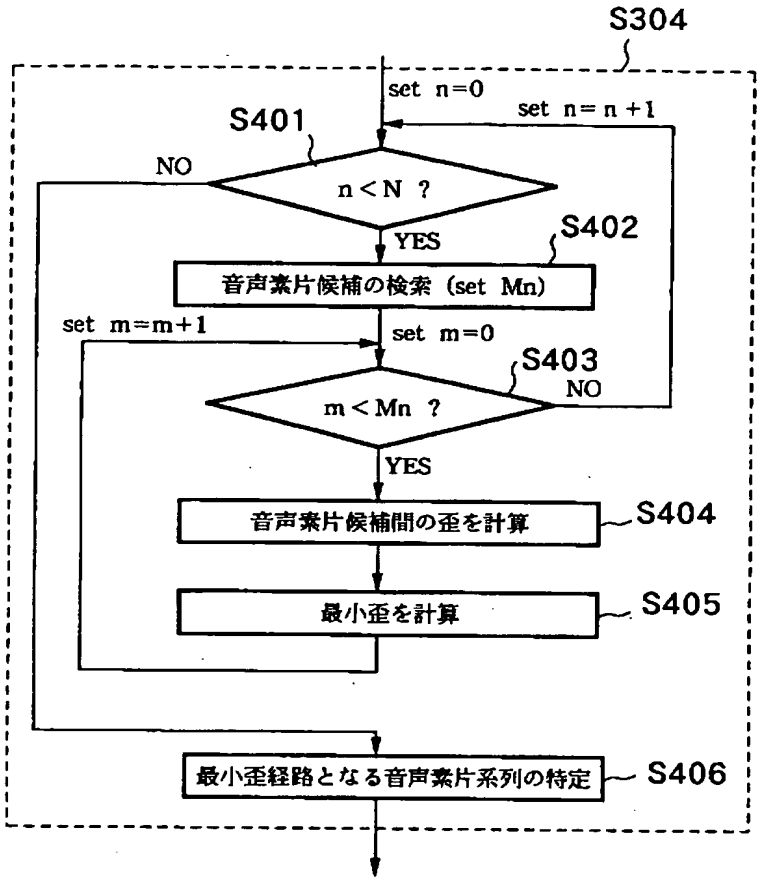
【図２】


テキスト入力部
201
言語解析部
203
解析辞書
202
韻律生成規則保持部
204
韻律生成部
205
音声素片保持部
206
音声素片選択部
207
音声波形出力部
209
音声素片編集・接続部
208

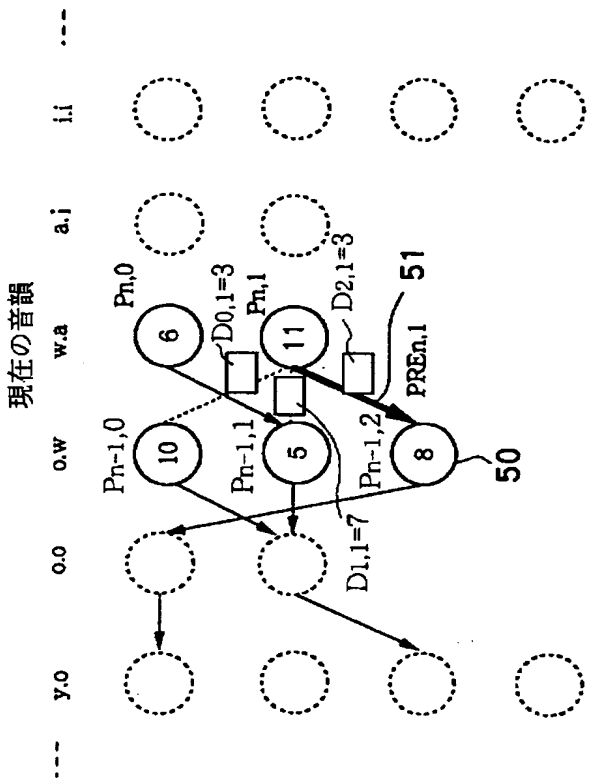
【図３】


スタート
テキスト入力
S301
言語解析
S302
韻律生成
S303
音声素片選択
S304
音声素片編集・接続
S305
音声波形出力
S306
エンド

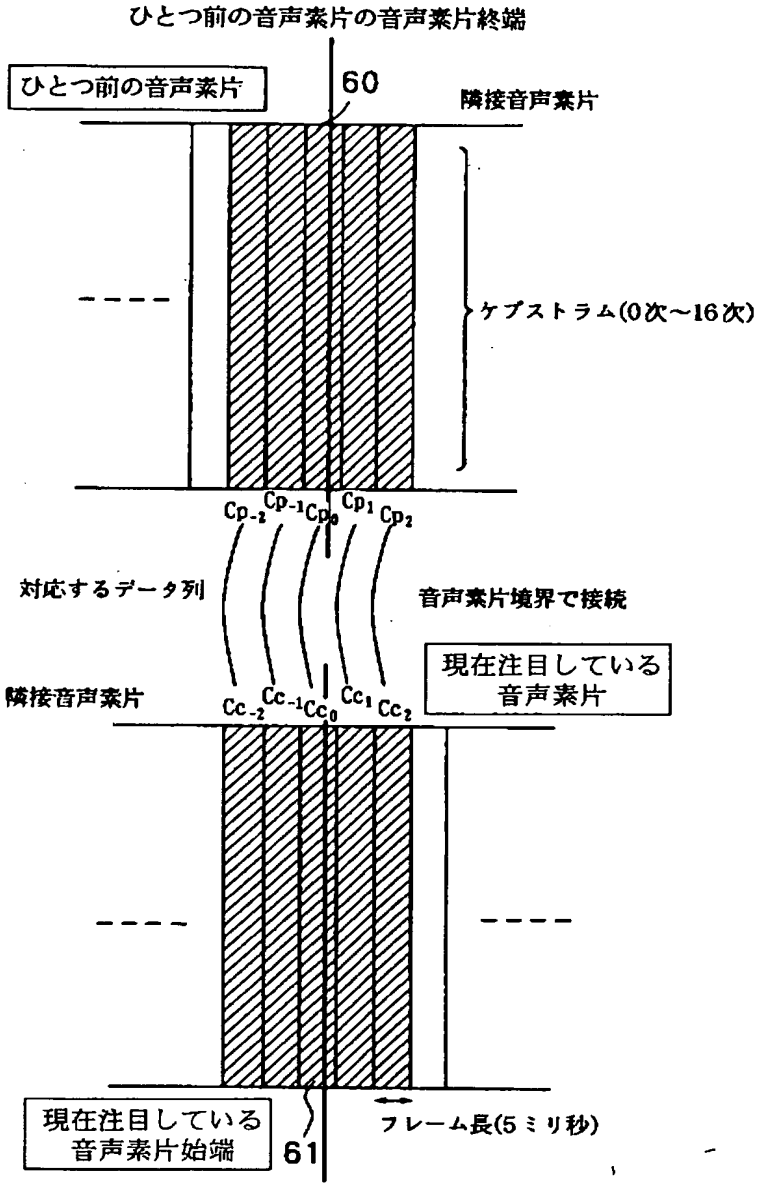
【図４】


S304
set n=0
set n=n+1
S401
NO
n<N ?
YES
S402
音声素片候補の検索 (set Mn)
set m=m+1
set m=0
S403
NO
m<Mn ?
YES
音声素片候補間の歪を計算
S404
最小歪を計算
S405
最小歪経路となる音声素片系列の特定
S406

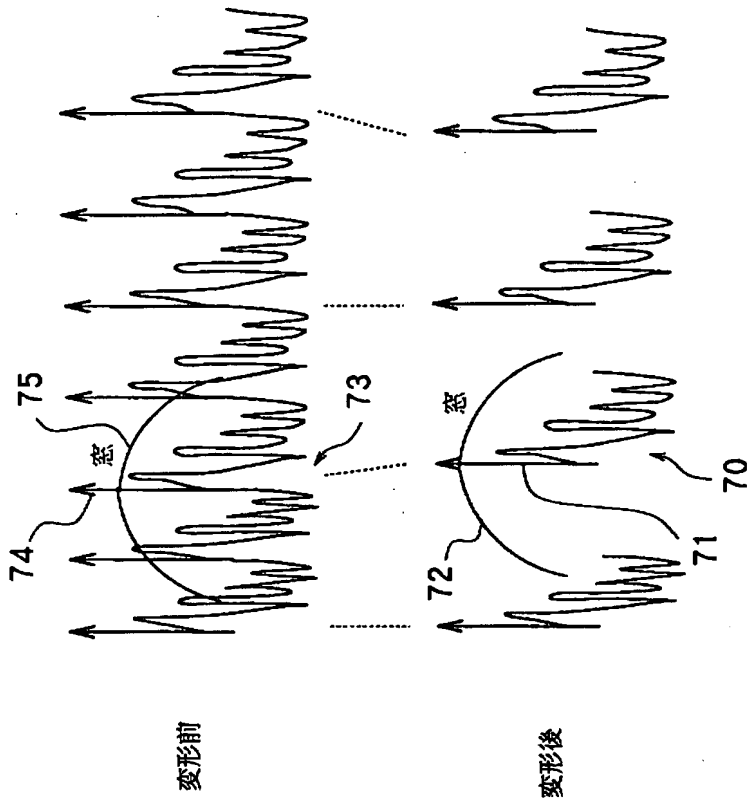
【図５】


現在の音韻
--- y.o o.o o.w w.a a.i i.i ---
Pn,0
Pn,1
Pn-1,0
Pn-1,1
Pn-1,2
D0,1=3
D2,1=3
D1,1=7
PREn,1
6
11
10
5
8
51
50

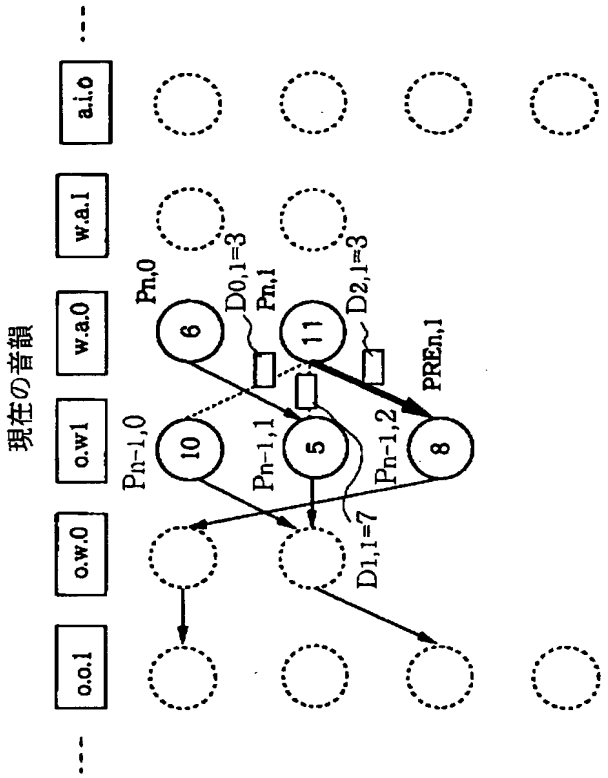
【図６】


ひとつ前の音声素片の音声素片終端
ひとつ前の音声素片
60
隣接音声素片
ケプストラム(0次～16次)
Cp-2 Cp-1 Cp0 Cp1 Cp2
対応するデータ列
音声素片境界で接続
現在注目している
音声素片
隣接音声素片
Cc-2 Cc-1 Cc0 Cc1 Cc2
現在注目している
音声素片始端
61
フレーム長(5ミリ秒)

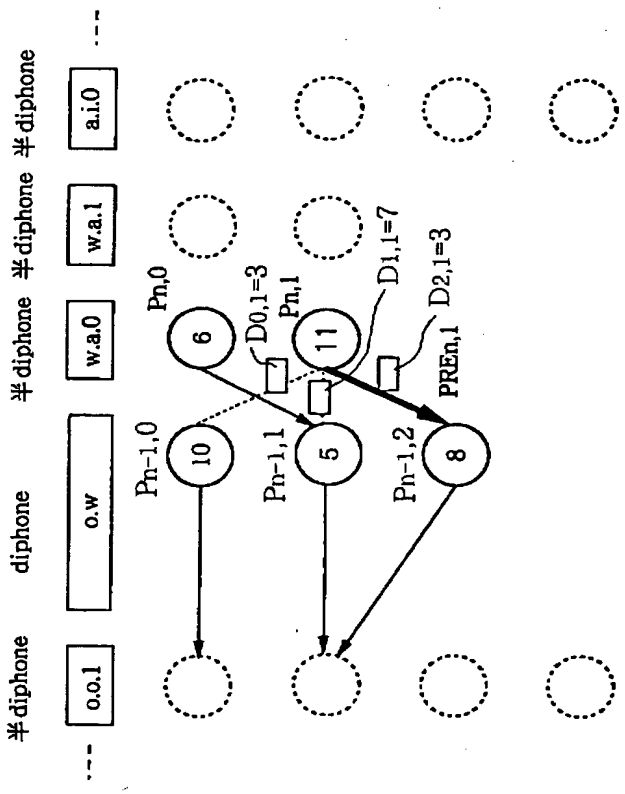
【図７】



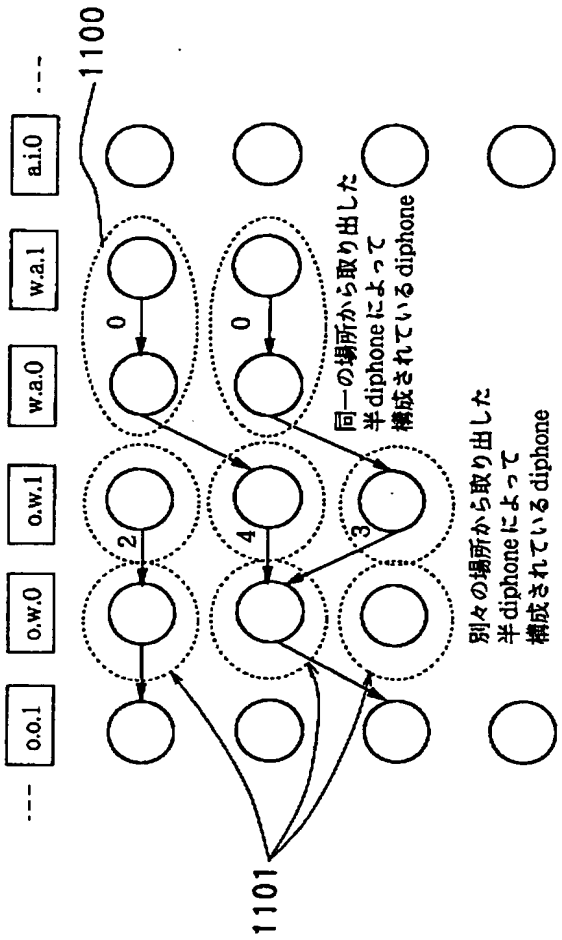
【図８】


現在の音韻
a.i.0
w.a.1
w.a.0
o.w.1
o.w.0
o.o.1
Pn,0
Pn,1
PREn,1
Pn-1,0
Pn-1,1
Pn-1,2
D0,1=3
D2,1=3
D1,1=7
6
11
10
5
8

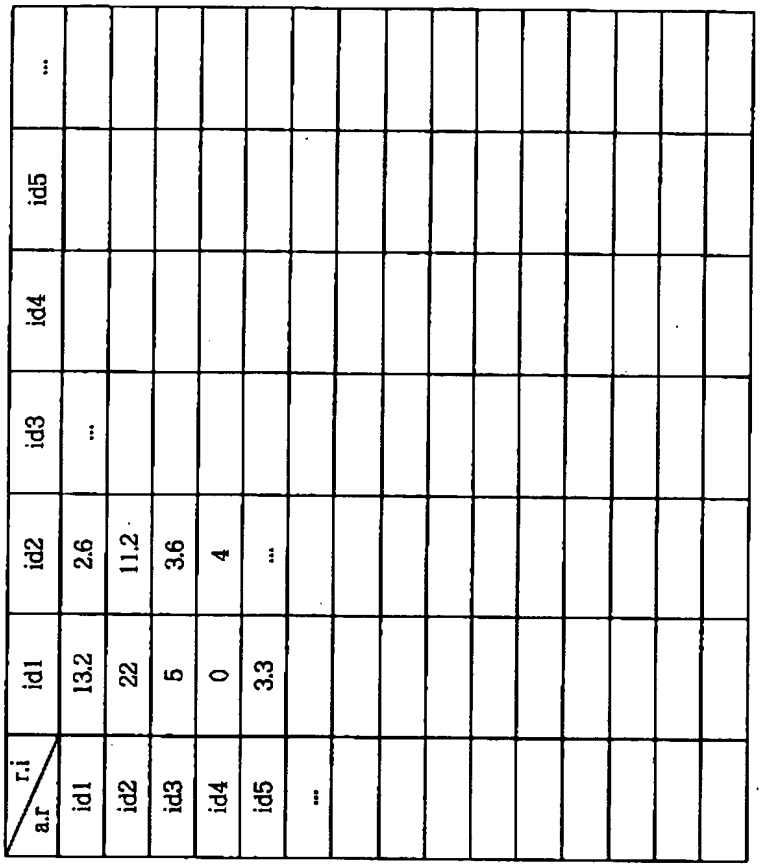
【図９】


半diphone
diphone
半diphone
半diphone
半diphone
o.o.1
o.w
w.a.0
w.a.1
a.i.0
Pn,0
Pn,1
Pn-1,0
Pn-1,1
Pn-1,2
D0,1=3
D1,1=7
D2,1=3
PREn,1
6
11
10
5
8

【図１０】


1100
a.i.0
w.a.1
w.a.0
o.w.1
o.w.0
o.o.1
0
0
2
4
3
同一の場所から取り出した
半diphoneによって
構成されている diphone
別々の場所から取り出した
半diphoneによって
構成されている diphone
1101

【図１１】



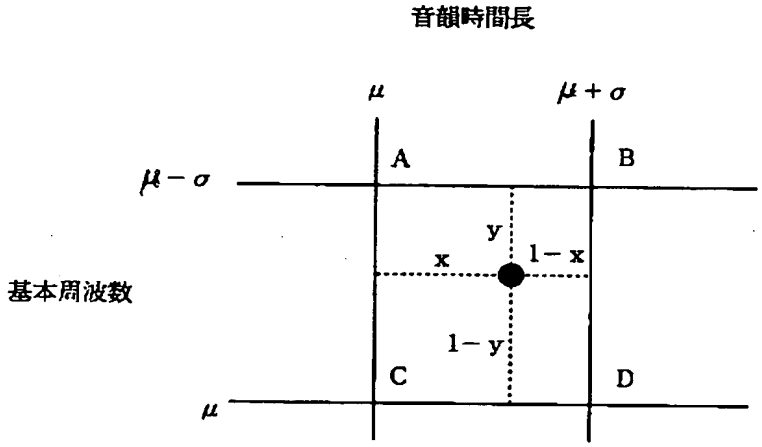
r.i / a.r	id1	id2	id3	id4	id5	...
id1	13.2	2.6	...			
id2	22	11.2				
id3	5	3.6				
id4	0	4				
id5	3.3	...				
...						

【図１２】

音韻時間長

基本周波数	$\mu-2\sigma$	$\mu-\sigma$	μ	$\mu+\sigma$	$\mu+2\sigma$
$\mu-2\sigma$	26	25	15	13	14
$\mu-\sigma$	20	24	19	18	17
μ	15	19	10	13	20
$\mu+\sigma$	19	22	16	22	26
$\mu+2\sigma$	25	27	26	30	31

【図１３】



【書類名】　要約書

【要約】

【課題】　接続歪と変形歪に基づく歪が小さくなるように音声素片を選択して、音声合成の音質劣化を抑制する。

【解決手段】　所定の音韻環境に対応付けて複数の音声素片を保持する音声素片保持部２０６から音韻環境に対応する複数の音声素片を抽出し（Ｓ４０２）、それら抽出された複数の音声素片のそれぞれの歪を算出し（Ｓ４０４）、音韻環境に基づいて決定される所定区間内で最小歪を求め（Ｓ４０５）、最小歪経路となる音声素片列を選択し（Ｓ４０６）、その音声素片を編集・接続して音声合成を行なう。

【選択図】　図４

出　願　人　履　歴　情　報

識別番号　　〔０００００１００７〕

１．変更年月日　　１９９０年　８月３０日
［変更理由］　新規登録
住　所　東京都大田区下丸子３丁目３０番２号
氏　名　キヤノン株式会社